

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

- 1 1. (currently amended) A method of focused crawling, comprising:
2 accessing a query input, the query input including at least a first query part and a second
3 query part;
4 crawling a plurality of documents, at least some of the plurality of documents including
5 links to each other, the crawling at least partly guided by a crawl metric, the crawl metric at least
6 partly determined by a mechanism and by the first query part; and
7 returning target documents, the target documents being relevant to the second query part,
8 the target documents found from the plurality of crawled documents, the target documents
9 returned at least partly based on a search metric, the search metric at least partly determined by
10 the mechanism and by the second query ~~part~~ part;
11 wherein the crawl metric comprises a metric that quantifies priority for crawling links
12 emanating from a certain document within context of the focused crawling and
13 wherein the search metric comprises a metric that quantifies relevance or importance of a
14 document to the query input.
- 1 2. (previously amended) The method of claim 1, wherein relevance includes importance.
- 1 3. (previously amended) A method of focused crawling, comprising:
2 accessing a query input including at least a first query part and a second query part;
3 crawling a plurality of documents, at least some of the plurality of documents including
4 links to each other, the crawling at least partly guided by a crawl metric, the crawl metric at least
5 partly determined by a first mechanism and by the first query part; and
6 returning target documents, the target documents being relevant to the second query part,
7 the target documents found from the plurality of crawled documents, the target documents

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

8 returned at least partly based on a search metric, the search metric at least partly determined by a
9 second mechanism and by the second query part.

1 4. (previously amended) The method of claim 3, wherein relevance includes importance.

1 5. (previously amended) A method of focused crawling, comprising:

2 accessing a query input;

3 crawling a plurality of documents, the documents including links to each other, and the

4 crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a

5 first mechanism, the first mechanism including a first combination, the first combination

6 including a first plurality of one or more procedures, the first plurality of one or more procedures

7 including one or more of: 1) evaluating relevance of documents using logical expressions of

8 keywords and phrases, 2) evaluating relevance of documents using a template including a

9 plurality of one or more template portions, at least one of the template portions including a first

10 plurality of one or more hierarchical levels, 3) evaluating relevance of documents using a link

11 structure of the crawled documents, and 4) evaluating relevance based on freshness of

12 documents; and

13 returning target documents, the target documents being relevant to the query input, the

14 target documents found from the plurality of crawled documents, the target documents returned

15 at least partly based on a search metric, the search metric at least partly determined by a second

16 mechanism, the second mechanism including a second combination, the second combination

17 being different from the first combination, the second combination including a second plurality

18 of one or more procedures, the second plurality of procedures including one or more of: 1)

19 evaluating relevance of documents using logical expressions of keywords and phrases, 2)

20 evaluating relevance of documents using a template including a plurality of one or more template

21 portions, at least one of the template portions including a second plurality of one or more

22 hierarchical levels, 3) evaluating relevance of documents using a link structure of the crawled

23 documents, and 4) evaluating relevance based on freshness of documents.

1 6. (previously amended) The method of claim 5, wherein relevance includes importance.

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

1 7. (currently amended) The method of claim 5, wherein at least one of the first mechanism
2 and the second mechanism includes:
3 associating a weight to each of the evaluated relevance ~~relevances~~ of the procedures; and
4 combining the evaluated relevance ~~relevances~~ and the weights of the relevance
5 ~~relevances~~.

1 8. (previously amended) The method of claim 5, wherein one or more of: 1) the first
2 plurality of one or more hierarchical levels and 2) the second plurality of one or more
3 hierarchical levels, includes at least one or more heading levels and one or more content levels.

1 9. (previously amended) The method of claim 5, wherein evaluating relevance includes
2 evaluating relevance of at least a first document and one or more of a first plurality of one or
3 more referring documents and a second plurality of one or more referring documents, each of the
4 first plurality of one or more referring documents referring to the first document directly, and
5 each of the second plurality of referring documents referring to the first document indirectly
6 through one or more documents.

1 10. (previously amended) A method of focused crawling, comprising:
2 accessing a query input;
3 crawling a plurality of documents, the documents including links to each other, and the
4 crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a
5 first mechanism, the first mechanism including a first combination, the first combination
6 including a first plurality of one or more procedures, the first plurality of one or more procedures
7 including one or more of: 1) evaluating relevance of documents using logical expressions of
8 keywords and phrases, 2) evaluating relevance of documents using a template including a
9 plurality of one or more template portions, at least one of the template portions including a first
10 plurality of one or more hierarchical levels, 3) evaluating relevance of documents using a link
11 structure of the crawled documents, and 4) evaluating relevance based on freshness of
12 documents; and

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

13 returning target documents, the target documents being relevant to the query input, the
14 target documents found from the plurality of crawled documents, the target documents returned
15 at least partly based on a search metric, the search metric at least partly determined by a second
16 mechanism, the second mechanism including a second combination, the second combination
17 being different from the first combination, the second combination including a second plurality
18 of one or more procedures, the second plurality of procedures including one or more of: 1)
19 evaluating relevance of documents using logical expressions of keywords and phrases, 2)
20 evaluating relevance of documents using a template including a plurality of one or more template
21 portions, at least one of the template portions including a second plurality of one or more
22 hierarchical levels, 3) evaluating relevance of documents using a link structure of the crawled
23 documents, and 4) evaluating relevance based on freshness of documents,

24 wherein the procedure, of the first plurality of one or more procedures, of evaluating
25 relevance of documents using a link structure of the crawled documents, includes:

26 accessing a first plurality of documents from a database of a plurality of received
27 documents, the plurality of received documents including crawled documents, the first plurality
28 of documents to be ranked;

29 generating a graph of the first plurality of documents;

30 assigning weights to one or more nodes of the graph;

31 finding an assignment of weights to one or more nodes of the graph, by
32 propagating weights through the graph, the assignment of weight to a node based at least in part
33 on calculating a weighted sum of weights propagated from neighboring nodes; and

34 generating a ranked list of at least the first plurality of documents, the ranked list
35 at least partly generated from the graph.

1 11. (previously amended) The method of claim 10, wherein relevance includes importance.

1 12. (currently amended) The method of claim 10, wherein at least one of the first mechanism
2 and the second mechanism includes:

3 associating a weight to each of the evaluated relevance ~~relevances~~ of the procedures; and

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

4 combining the evaluated relevance ~~relevances~~ and the weights of the evaluated relevance
5 ~~relevances~~.

1 13. (previously amended) The method of claim 10, wherein one or more of: 1) the first
2 plurality of one or more hierarchical levels and 2) the second plurality of one or more
3 hierarchical levels, includes at least one or more heading levels and one or more content levels.

1 14. (previously amended) The method of claim 10, wherein evaluating relevance includes
2 evaluating relevance of at least a first document and one or more of a first plurality of one or
3 more referring documents and a second plurality of one or more referring documents, each of the
4 first plurality of one or more referring documents referring to the first document directly, and
5 each of the second plurality of referring documents referring to the first document indirectly
6 through one or more documents.

1 15. (previously amended) The method of claim 10, wherein the procedure, of the first
2 plurality of one or more procedures, of evaluating relevance of documents using a link structure
3 of the crawled documents, further comprises:

4 expanding the graph with a second plurality of one or more documents from the database,
5 wherein a third plurality includes a union of the first plurality of documents and the second
6 plurality of documents, and the third plurality of documents is smaller than the plurality of
7 received documents.

1 16. (previously amended) The method of claim 10, wherein the procedure, of the first
2 plurality of one or more procedures, of evaluating relevance of documents using a link structure
3 of the crawled documents, further comprises:

4 expanding the graph with a second plurality of one or more documents from the database,
5 such that a third plurality includes a union of the first plurality of documents and the second
6 plurality of documents, and the third plurality of documents is smaller than the plurality of
7 received documents, the second plurality including one or more of: 1) one or more documents
8 connected within a first specified number of links in a forward direction from one or more
9 documents of the first plurality of documents, the forward direction being forward from the first

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

10 plurality of documents, and 2) one or more documents connected within a second specified
11 number of links in a backward direction from one or more documents of the first plurality of
12 documents, the backward direction being backward from the first plurality of documents.

1 17. (previously amended) The method of claim 10, wherein the procedure, of the first
2 plurality of one or more procedures, of evaluating relevance of documents using a link structure
3 of the crawled documents, further comprises:
4 expanding the graph with a second plurality of one or more documents from the database,
5 such that a third plurality includes a union of the first plurality of documents and the second
6 plurality of documents, and the third plurality of documents is smaller than the plurality of
7 received documents, the second plurality including one or more of: 1) all documents connected
8 within a first specified number of links in a forward direction from one or more documents of the
9 first plurality of documents, the forward direction being forward from the first plurality of
10 documents, and 2) all documents connected within a second specified number of links in a
11 backward direction from one or more documents of the first plurality of documents, the
12 backward direction being backward from the first plurality of documents.

1 18. (previously amended) The method of claim 10, wherein the first plurality of documents
2 includes recently received documents of the plurality of received documents.

1 19. (previously amended) The method of claim 10, wherein the procedure, of the first
2 plurality of one or more procedures, of evaluating relevance of documents using a link structure
3 of the crawled documents, further comprises:
4 shrinking the graph by removing one or more nodes of the graph.

1 20. (previously amended) The method of claim 10, wherein the procedure, of the first
2 plurality of one or more procedures, of evaluating relevance of documents using a link structure
3 of the crawled documents, further comprises:
4 shrinking the graph by combining one or more sets of one or more nodes of the graph.

5 21. (previously presented) The method of claim 20, wherein the combining is based on
6 common characteristics of the nodes or relationships between the nodes.

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

1 22. (previously presented) The method of claim 10, wherein the propagating weights
2 through the graph occurs up to a limited node distance.

1 23. (previously presented) The method of claim 10, wherein weights assigned to a document
2 include at least one of relevance of the document to the query input and importance of the
3 document independent of the query input.

1 24. (canceled) The method of claim 10, wherein relevance includes importance.